CHARTER MARK

The Patent Office

PCT/GB 99/00816

17 MARCH 1999

INVESTOR IN PEOPLE

GB99/816

The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH

REC'D 12 MAY 1999
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation and Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words public limited company may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Signed [signature]

Dated 20 April 1999

Patents Form 1/77

Patents Act 1977
(Rule 16)

17 MAR 1998

The Patent Office

9
17MAR98 E346146-1 D02246
P01/7700 25.00 - 9805576.7

[redacted]equest for a grant of a patent

[redacted] notes on the back of this form. You can also get [redacted]anatory leaflet from the Patent Office to help [redacted]fill in this form)*

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1.	Your reference	P/4827.GB ATM
2.	Patent application number *(The Patent Office will fill in this part)*	17 MAR 1998 **9805576.7**
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	Medical Research Council 20 Park Crescent London W1N 4AL
	Patents ADP number *(if you know it)*	
	If the applicant is a corporate body, give the country/state of its incorporation	GB 0596007001
4.	Title of the invention	Proteins
5.	Name of your agent *(if you have one)*	D YOUNG & CO
	"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	21 NEW FETTER LANE LONDON EC4A 1DA
	Patents ADP number *(if you have one)*	59006
6.	If you are declaring priority from one or more earlier patent applications, give the country and date of filing of the or each of these earlier applications and (if you know it) the or each application number	Country / Priority application number *(if you know it)* / Date of filing *(day/month/year)*
7.	If this application is divided or otherwise derived from an earlier UK application, give the number and filing date of the earlier application	Number of earlier application / Date of filing *(day/month/year)*

8. Is a statement of inventorship and [redacted]ght to grant of a patent required in support of this request? (*Answer 'Yes' if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

YES

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	0
Description	10
Claims*(s)*	0
Abstract	0
Drawing*(s)*	4

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*

Request for preliminary examination and search *(Patents Form 9/77)*

Request for substantive examination *(Patents Form 10/77)*

Any other documents *(please specify)*

11. I/We request the grant of a patent on the basis of this application.

Signature

Date

17 03 98

D YOUNG & CO
Agents for the Applicants

12. Name and daytime telephone number of the person to contact in the United Kingdom

Dr A Maschio 01703 634816

Warning
After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes
a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505

b) Write your answers in capital letters using black ink or you may type them

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Recognition of DNA methylation by zinc fingers

Yen Choo

Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, U.K.

Zinc fingers are small DNA-binding motifs which occur in a large family of eukaryotic transcription factors [1, 2]. The DNA-binding specificity of zinc fingers can be altered by protein engineering, for instance using phage display [3], to create novel protein domains which recognise predetermined sequences. It has been proposed that tailored DNA-binding domains of this type can be incorporated into proteins such as restriction enzymes [4] and transcription factors [5-8], in order to target particular DNA sequences or genes. The zinc finger domains studied so far - whether naturally occurring, designed or selected - can bind specifically to various DNA sites containing the four major DNA bases: A, G, C and T. However, the DNA of many organisms also contains a fifth base, 5-methylcytosine (5-meC), which arises from specific methylation of cytosine, and which is used to mark the genome or to increase its information content [9]. 5-methylcytosine is well known to affect protein-DNA interactions [10], for instance inhibiting cleavage of DNA by certain restriction enzymes. In vertebrates, cytosine is frequently methylated when directly preceding guanine, as in the dinucleotide CpG [11]. This type of methylation generally down-regulates vertebrate gene expression, and can also prevent the binding of many eukaryotic transcription factors to DNA [11, 12]. Yet the zinc finger transcription factors tested to date, Sp1 and YY1, are not affected by CpG methylation of their DNA binding sites [13, 14], suggesting that zinc fingers are incapable of discriminating between cytosine and 5-meC. On the contrary, it is shown below, that phage-selected zinc finger DNA-binding domains can distinguish the two closely related bases in the context of the CpG dinucleotide and are hence capable of differential binding to their DNA sites depending on the methylation status of cytosine.

A phage display library of the three-finger DNA-binding domain from the mouse transcription factor Zif268 has been described [15], in which finger 2 was randomised in those positions of the DNA-recognition helix which were thought to function in DNA binding (Fig. 1a). The library was screened using a version of the Zif268 DNA binding site, GCG<u>TGG</u>GCG, in which the triplet bound by the

randomised middle finger (underlined) was replaced by a given target sequence. When the sequences GCGGCGGCG and GCGGTGGCG were used in selections, some zinc finger DNA-binding domains were selected which bound both sequences equally well (Fig. 1b, c) [15, 16]. However, two additional zinc finger families were isolated which were capable of differential binding to the two closely related sites (Fig. 1b, c) [15, 16]. Sequence-specific recognition required discrimination of the central base in the binding site by amino acids in position 3 of the recognition helix of the selected zinc fingers, and it was noted that aspartate was selected to bind opposite cytosine in the triplet GCG, while alanine was selected opposite thymine in the triplet GTG. The correlation between thymine and alanine was particularly significant, as it implied a van der Waals interaction between the amino acid side-chain and the 5-methyl group of the base [17]. Indeed, when thymine was mutated to deoxyuracil in the binding sites of such fingers there was a dramatic decrease in the strength of the intermolecular interaction (Fig 1c). This showed that these zinc fingers were capable of specifically recognising a 5-methyl group, and suggested that similar fingers might be selected which bind 5-meC by the same token. In order to determine whether this was indeed possible, the phage display library was screened with the synthetic binding site GCGGMGGCG, containing a 5-meC base analogue (M). After 5 rounds of selection, zinc finger phage were tested for binding to 5-meC and cytosine in the context of the above site, and those capable of specifically binding the methylated site were sequenced in the region of the zinc finger gene. Two different clones were isolated, which were identical to the DNA-binding domains previously selected using the binding site GCGGTGGCG.

Hence the various zinc finger phage selections described above yielded different fingers able to bind the generic DNA sequence GCGGNGGCG, where N was either thymine, cytosine or 5-meC. A full complement of fingers was selected for recognition of the cytosine/5-meC pair in the above context, some of which recognised one type of base exclusively, while others bound both bases equally well (Figures 1c and 2). However, any fingers which recognised 5-meC were unable to discriminate against thymine (Fig. 1c).

The zinc finger amino acid residues which were selected by the interaction between the randomised recognition helix and the central base of the DNA binding site could be rationalised in terms of previously elucidated zinc finger-DNA recognition rules [18]. Fingers with alanine in position 3 of the recognition helix specifically bound 5-meC and thymine owing to a tight hydrophobic interaction between the side chain and the 5-methyl group which is present in both bases. In contrast, a finger with valine in position 3 was also able to accommodate cytosine in addition to the two methylated bases, presumably by the use of different rotamers. Fingers with aspartate in position 3 bound cytosine specifically, perhaps by forming a ring structure which packs against the pyrimidine as was observed in the refined crystal structure of Zif268 [19], although it is noted that in this case deoxyuracil, which is bound only weakly, might have been expected to be equally acceptable (Fig. 1c).

Zinc fingers were selected from a phage display library to bind the minor base 5-meC, but not cytosine, and thus to discriminate the methylation status of cytosine in the context of a CpG dinucleotide. Since this DNA-binding motif is capable of discriminating 5-meC, the zinc finger transcription factor Sp1 may have evolved to recognise its DNA binding sites regardless of their methylation status. On the other hand, more specialised zinc finger transcription factors may respond to methylation of their binding sites in the same way demonstrated for phage-selected zinc finger DNA-binding domains. Moreover, because the DNA binding differences shown in this paper resulted from recognition of only a single methyl functional group, zinc fingers which specifically discriminate methylation of multiple CpG dinucleotides in their binding sites should be capable of a more pronounced response to methylation of DNA.

Recently, further zinc fingers which bind methylated DNA have been selected from phage display libraries, and discrimination between cytosine and 5-meC has also been rationally engineered into tailored zinc finger domains (unpublished). Zinc finger DNA-binding domains designed to discriminate 5-meC in predetermined sequences are potentially powerful tools which may find applications in the study and manipulation of

DNA methylation, mutagenesis, restriction/modification systems, gene dosage compensation, genomic imprinting, and the control of gene expression.

FIGURE LEGENDS

Fig 1a Alignment of the amino acid sequence of the three fingers from Zif268 used in the phage display library. Randomised residue positions in the α-helix of finger 2 are marked 'X' and are numbered above the alignment relative to the first helical residue (position 1). Residues which form the hydrophobic core are circled; zinc ligands are written as white letters on a black circle background; and positions comprising the secondary structure elements of a zinc finger are marked below the sequence.

Fig 1b Amino acid sequences of the variant α–helical regions from some zinc fingers selected by phage display using the DNA binding site GCGGNGGCG where the central (bold) nucleotide of the middle (underlined) triplet was either: (i) 5-methylcytosine, (ii) thymine, or (iii) cytosine. Amino acid sequences are listed below the DNA oligo used in their selection. Amino acid positions are numbered above the aligned sequences relative to the first helical residue (position 1). Circled residues (in position 3) are predicted to contact the middle nucleotide of the binding site.

Fig 1c Phage ELISA binding assay showing discrimination of pyrimidines by representative phage-selected zinc fingers. The matrix shows three different zinc finger phage clones (x, y and z) reacted with four different DNA binding sites present at a concentration of 3nM. Binding is represented by vertical bars which indicate the OD obtained by ELISA [16]. The amino acid sequences of the variant α-helical regions from the selected zinc fingers were: REDVLIRHGK (x), RADALMVHKR (y), and RGPDLARHGR (z). The DNA sequences contained the generic binding site GCGGNGGCG, where the central (bold) nucleotide was either: uracil (U), thymine (T), cytosine (C), or 5-methylcytosine (M).

Fig. 2 Effect of cytosine methylation on DNA binding by phage-selected zinc fingers. Graphs show three different zinc finger phage binding to the DNA sequence

GCGGCGGCG in the presence (ed!! key: circle) and absence (ed!! key: triangle) of methylation of the central base (bold). The zinc finger clones tested contained variant α-helical regions of the middle finger as follows: (a) RADALMVHKR, (b) RGPDLARHGR and (c) REDVLIRHGK. These respective zinc finger clones preferentially bind their cognate DNA site in the presence, absence, or regardless of cytosine methylation.

REFERENCES

1. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J* 4, 1609-1614 (1985).

2. Pabo, C. O. & Sauer, R. T. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annual Review of Biochemistry* **61**, 1053-95 (1992).

3. Choo, Y. & Klug, A. Designing DNA-binding proteins on the surface of filamentous phage. *Curr. Opin. Biotech.* 6, 431-436 (1995).

4. Kim, Y.-G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to *Fok* I cleavage domain. *Proc. Natl. Acad. Sci. USA* **93**, 1156-1160 (1996).

5. Choo, Y., Sanchez-Garcia, I. & Klug, A. *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**, 642-645 (1994).

6. Kim, J.-S., Kim, J., Cepek, K. L., Sharp, P. A. & Pabo, C. O. Design of TATA box-binding protein/zinc finger fusions for targeted regulation of gene expression. *Proc Natl Acad Sci USA* 94, 3616-3620 (1997).

7. Liu, Q., Segal, D. J., Ghiara, J. B. & Barbas III, C. F. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci USA* **94**, 5525-5530 (1997).

8. Choo, Y., Castellanos, A., Garcia-Hernandez, B., Sanchez-Garcia, I. & Klug, A. Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. *J. Mol. Biol* 273, 525-532 (1997).

9. *DNA methylation: molecular Biology and biological significance* (Jost, J. P. & Saluz, H. P., Eds.), Birkhauser Verlag, Basel. (1993).

10. Adams, L. P. DNA methylation. *Biochem. J.* **265**, 309-320 (1990).

11. Jones, P. A., Rideout III, W. M., Shen, J.-C., Spruck, C. H. & Tsai, Y. C. Methylation, mutation and cancer. *BioEssays* **14**, 33-36 (1992).

12. Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* 3, 226-231 (1993).

13. Holler, M., Westin, G., Jiricny, J. & Schaffner, W. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev* **2**, 1127-1135 (1988).

14. Gaston, K. & Fried, M. CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. *Nucleic Acids Res* **23**, 901-909 (1995).

15. Choo, Y. & Klug, A. Toward a code for the interactions of zinc fingers with DNA: Selection of randomised zinc fingers displayed on phage. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11163-11167 (1994).

16. Choo, Y. & Klug, A. Selection of DNA binding sites for zinc fingers using rationally randomised DNA reveals coded interactions. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11168-11172 (1994).

17. Ivarie, R. Thymine methyls and DNA-protein interactions. *Nucleic Acids Res.* **15**, 9975-9983 (1987).

18. Choo, Y. & Klug, A. Physical basis of a protein-DNA recognition code. *Curr. Opin. Str. Biol.* **7**, 117-125 (1997).

19. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. Zif268 protein-DNA complex refined at 1.6A: a model system for understanding zinc finger interactions. *Structure* **4**, 1171-1180 (1996).

ACKNOWLEDGEMENT

I would like to thank Aaron Klug for his help and support in this research.

```
              -1123456789
MAEERPYACPVESCDRRFSRSDELTRHIRIHT
  GQKPFQCRI--CMRNFSXXXXLXXHX{R/K}THT
  GEKPFACDI--CGRKFARSDERKRHTKIHLRQKD
      β        β        α
```

Choo Fig 1A

(i) GCGG**M**GGCG

-1123456789
RADALMVHKR
RGDALTSHER

(ii) GCGG**T**GGCG

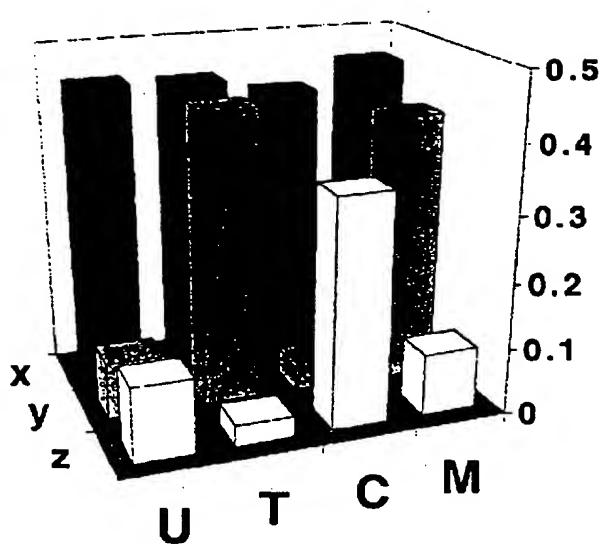
-1123456789
RADALMVHKR
RGDALTSHER
RVDALEAHRR
REDVLIRHGK

(iii) GCGG**C**GGCG

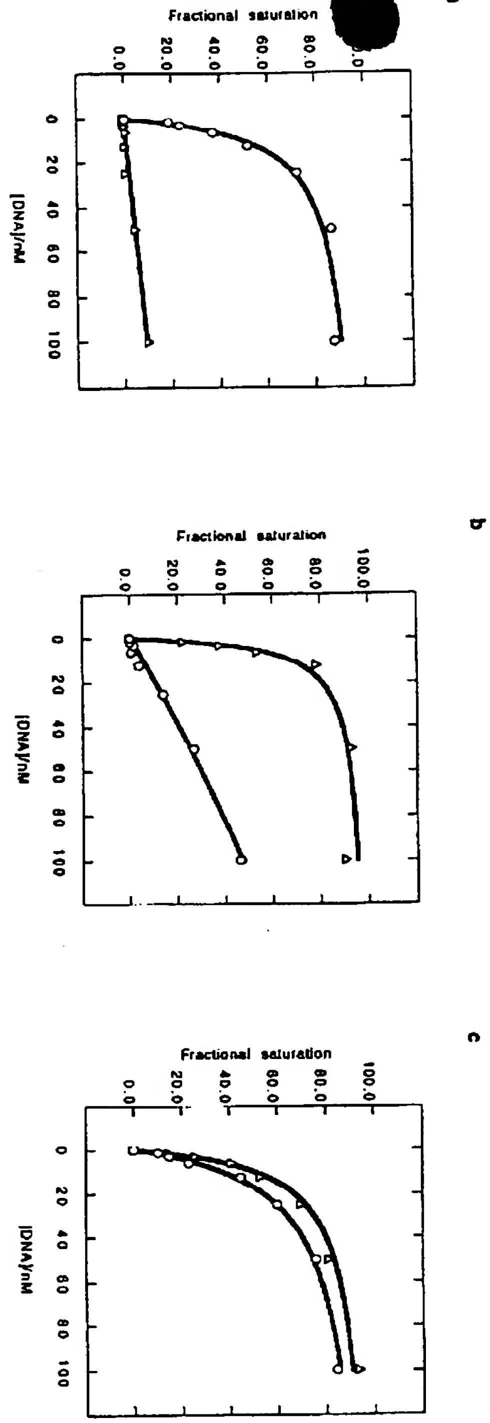
-1123456789
RGPDLARHGR
REDVLIRHGK

Choo FIG 1B

Worksheet2 Chart 3



Chao Fig 1C


Fractional saturation
0.0
20.0
40.0
60.0
80.0
[DNA]/nM
0
20
40
60
80
100
b
Fractional saturation
0.0
20.0
40.0
60.0
80.0
100.0
[DNA]/nM
0
20
40
60
80
100
c
Fractional saturation
0.0
20.0
40.0
60.0
80.0
100.0
[DNA]/nM
0
20
40
60
80
100

Chao Fig 2